PERGAMON Computers and Mathematics with Applications 37 (1999) 9–17

An International Journal
computers & mathematics with applications

Some Old Traditions in Mathematics and in Mathematical Education

E. A. GALPERIN
Département de Mathématiques, Université du Québec à Montréal
C.P. 8888, Succ. Centre-Ville, Montréal, Québec, Canada H3C 3P8

Abstract—Continuing the discussion by Azbelev about the harmful influence of some classical traditions on the study of differential equations [1], this note presents further instances of old traditions and of their impact on the development of certain areas of mathematics and its applications.

Keywords—Classical traditions, Mathematical education, Derivatives, Optimality, Noncausality.

1. PHYSICAL NONEQUIVALENCE OF RIGHT AND LEFT DERIVATIVES OF CONTINUOUSLY DIFFERENTIABLE FUNCTIONS

Consider a scalar continuously differentiable function $x(t)$. Recall related mathematical definitions.

DEFINITION 1. *A function $x(t)$ is right differentiable at a point $t = t_0$ if there exists a limit*

$$\dot{x}\left(t_0^+\right) = \frac{dx}{dt}\bigg|_{t_0^+} = \lim_{\Delta t \to 0} \frac{x(t+\Delta t) - x(t)}{\Delta t}\bigg|_{t=t_0, \Delta t>0}. \tag{1.1}$$

This limit is called *right derivative* of $x(t)$ at t_0.

A function $x(t)$ is *left differentiable* at a point $t = t_0$ if there exists a limit $\dot{x}(t_0^-)$ expressed by the same formula (1.1), but with $\Delta t < 0$; this limit is called *left derivative* of $x(t)$ at t_0.

A function $x(t)$ is *differentiable* at a point $t = t_0$ if there exist both $\dot{x}(t_0^+)$ and $\dot{x}(t_0^-)$, and $\dot{x}(t_0^+) = \dot{x}(t_0^-)$. This common value is denoted $\dot{x}(t_0)$ and called *derivative* of $x(t)$ at t_0.

DEFINITION 2. *A function $x(t)$ is continuously right differentiable at t_0 if there exists a neighborhood $\mathcal{N}_+[t_0, \delta) = [t_0, t_0 + \delta)$, $\delta > 0$, such that for every $t \in [t_0, t_0 + \delta)$, there exists right derivative $\dot{x}(t^+)$ which is continuous at t_0,*

$$\lim_{t \to t_0+0} \dot{x}\left(t^+\right) = \dot{x}\left(t_0^+\right). \tag{1.2}$$

A function $x(t)$ is continuously left differentiable at t_0 if there exists a neighborhood $\mathcal{N}_-[t_0, \delta) = (t_0 - \delta, t_0]$, $\delta > 0$, such that for every $t \in (t_0 - \delta, t_0]$, there exists left derivative $\dot{x}(t^-)$ which is continuous at t_0,

$$\lim_{t \to t_0-0} \dot{x}(t^-) = \dot{x}(t_0^-). \tag{1.3}$$

0898-1221/99/$ - see front matter Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX
PII: S0898-1221(99)00055-3

Finally, a function $x(t)$ is continuously differentiable at t_0 if there exist $\dot{x}(t^+)$ and $\dot{x}(t^-)$ in respective neighborhoods, both continuous at t_0, and $\dot{x}(t_0^+) = \dot{x}(t_0^-)$. If $\dot{x}(t^+)$ and $\dot{x}(t^-)$ are continuous not just at t_0, but at every $t \in (t_0-\delta, t_0+\delta)$, $\delta > 0$, then a function $x(t)$ is continuously differentiable in a neighborhood $(t_0 - \delta, t_0 + \delta)$ with the common value $\dot{x}(t) = \dot{x}(t^+) = \dot{x}(t^-)$ called derivative (continuous) of a function $x(t)$ in that neighborhood.

Almost all physical processes (not necessarily models thereof) present functions continuously differentiable almost everywhere (i.e., except some isolated points; or sets of measure zero, strictly speaking). The existence of a continuous one-sided derivative on an open interval implies continuous differentiability on that interval (see [2, Lemma 2.1, p. 176]). It is worth noting that mere finiteness of a one-sided derivative on an open interval implies differentiability almost everywhere on that interval [3]. Thus, from the mathematical point of view, one can identify left and right derivatives in process equations. This has been the classical tradition for centuries—not to distinguish left and right derivatives; derivative $\dot{x}(t)$ actually written in equations was supposed to be the right derivative.

The deficiency of this tradition can be clearly demonstrated on many physical systems, and most profoundly it affects, probably, the classical mechanics. In relation to mechanical motion, a function $x(t)$ can be interpreted as distance x at time t, then velocity of motion $v(t) = \frac{dx}{dt} = \dot{x}(t)$ and the acceleration $a(t) = \frac{dv}{dt} = \frac{d^2x}{dt^2} = \ddot{x}(t)$.

If we interpret, however, the velocity of motion as right derivative, $v(t) = \dot{x}(t^+)$, see (1.1), then at each current moment t, it is unmeasurable and unobservable since the value $x(t + \Delta t)$ for $\Delta t > 0$ is not yet in existence at the moment t. How then, one may ask, can we see velocities on speedometers? Yes, we see it, but what we see is the left derivative. If we wanted to observe the right derivative, then it should have been supplied with a delay, since $\dot{x}(t^+)$ can be physically rendered at $t + \Delta t$ or later and not at the moment t. Only left derivatives can be observed and employed without delay. If, following the classical tradition, we identify left and right derivatives and write those left derivatives in process equations as right derivatives, then unintended strong restrictions are actually imposed, especially in control applications.

For a controlled motion of a solid with a constant mass, Newton's second law of motion can be written as follows:

$$a(t) = \ddot{x}(t) = \dot{v}(t) = f(t, x, v, u), \qquad t \geq t_0, \tag{1.4}$$

where u is a control action and $f(\cdot)$ is an external force (a motive force [4]) which can depend on time t, coordinate x, and velocity $v = \dot{x}(t)$, as well as on control $u(t, x, v)$ that can itself depend on the same variables. Derivatives $\dot{v}(t)$ or $\ddot{x}(t)$ are right derivatives that project the motion into future moments of time, given initial conditions $x(t_0), v(t_0)$. This motion can be computed, e.g., by Euler's formulae implied by (1.1),

$$v(t + \Delta t) = v(t) + f(\cdot)\Delta t, \quad x(t + \Delta t) = x(t) + v(t)\Delta t, \qquad t \geq t_0, \quad \Delta t > 0. \tag{1.5}$$

From (1.4),(1.5), it is clear that $f(\cdot)$ and $u(\cdot)$ cannot contain accelerations or higher-order right derivatives of $x(t)$. Indeed, if we took $u = g(\dot{v})$ with $\dot{v} = \dot{v}(t^+)$ as right derivative, then $f(\cdot)$ in (1.4) would be undefined at moment t, and in (1.5) the function $f(\cdot)$ would contain the term

$$\frac{v(t + \Delta t) - v(t)}{\Delta t},$$

approximating $\dot{v}(t^+)$, so that the Euler scheme would not work. For these reasons, the introduction of accelerations and higher-order derivatives in the right-hand side of (1.4) was considered tabu.

However, Lamb in [5, p. 168] considered equations of motion of a solid in an ideal liquid, see Section 124, equations (1) with reference to Kirchhoff and Thomson (1871), where forces of the fluid pressure linearly depended on the acceleration of the solid itself, (see [5, equations (2) and (3), pp. 168 and 169]). In this case, the right-hand side of (1.4) contained $\dot{v}(t)$ linearly and in such a way that (1.4) could be resolved for $\dot{v}(t)$ yielding a new equation with the sole accelerations $\dot{v}(t)$ on the left and effective force on the right, so that there were no difficulties with undefined advance terms or with applicability of the Euler scheme.

In fact, if we introduce into (1.4) the second- and higher-order derivatives as *left* derivatives, the way they are, and take into account that those left derivatives can be physically measured and computed on-line, we can use a control

$$u = u\left(t, x, \dot{x}^-, \ddot{x}^-, \dots, x^{(k)-}\right), \tag{1.6}$$

with all the left derivatives we need, this not causing any problem with physical validity of (1.4), nor with the application of (1.5), or other point propagation schemes (Runge-Kutta, etc). Of course, additional initial conditions should be introduced and the right-hand side of (1.4) with control (1.6) should not be mistaken for a field of external forces, see [2]. With consideration of effective forces, it can be demonstrated that such systems are in agreement with Newton's laws and with the parallelogram law for addition of forces [2,6]. It is worth noting that acceleration assisted control is already in use in engineering, and this term has gained recognition in the literature [7–16] though without theoretical justification. Clear understanding of the physical difference between right and left derivatives will remove the tabu and open new horizons with the use of higher-order derivatives in control of dynamical systems and physical processes.

2. NONCAUSAL EQUATIONS AND INEQUALITIES ARE LEGITIMATE PHYSICAL MODELS

The principle of causality means simply that any event is a consequence of a cause. In application to differential equation

$$\dot{x}(t) = f(t, x[h(t)]), \tag{2.1}$$

it implies that representation (2.1) is causal if and only if $h(t) \le t$. If $h(t) \equiv t$, then (2.1) is an ordinary differential equation. If $h(t) \le t$ and $h(t) \not\equiv t$, then (2.1) is a delay differential equation. The principle of causality intertwines with the feedback control principle. Indeed, the function $f(\cdot)$ in (2.1) can be interpreted as feedback if $h(t) \le t$.

If however, $h(t) > t$ (feedforward control or model), then equation (2.1) is noncausal since $x[h(t)]$ is nonexistent at the moment t, so that velocity $\dot{x}(t)$ is not determined by a natural cause at time t. In this case, a function $x(t)$ is sought in accordance with equation (2.1) in order to fulfill our desire to have forward value $x[h(t)]$ at the moment $h(t) > t$ of yet nonexistent future.

Physically, a distinction between the two situations can be specified by saying that with $h(t) \le t$, equation (2.1) describes a *process* or a *motion* driven by a "force" embodied in $f(t, x[h(t)])$ that really exists at each moment t so that (2.1) represents a physical reality that exists, or can be built or materialized in some way. In contrast, with $h(t) > t$, equation (2.1) represents a *model* of some hypothetical process or motion which complies with our demand to have the value $x[h(t)]$ at future moment $t_* = h(t) > t$. Such a model cannot be materialized in terms of the entries written in (2.1).

REMARK. Formally, if $h(t) > t$, one can denote $h(t) = \tau$, so that $\tau > t = h^{-1}(\tau)$. Now, writing (2.1) in the form

$$\dot{x}\left[h^{-1}(\tau)\right] = f\left[h^{-1}(\tau), x(\tau)\right], \tag{2.2}$$

and differentiating (2.2) with respect to τ, it may be possible to resolve the obtained new equation for $\dot{x}(\tau)$, converting the advance differential equation into a delay differential equation. Such an equation, however, would be much more complicated than (2.1) and with extraneous solutions. We do not consider such cases here.

Mathematically, if equation (2.1) is considered as a selector of functions $\{x(t)\}$ satisfying (2.1), there is no principal difference between the cases $h(t) \le t$ and $h(t) > t$ as concerns selection processes by setwise global optimization methods [17–26]. However, numerical methods based on pointwise propagation of the solution from some initial and/or boundary conditions (such methods as Euler, Runge-Kutta schemes for ODE; finite difference, finite elements, and other schemes for PDE) do not work in the case where $h(t) > t$. Indeed, for a functional differential equation (2.1) with $h(t) > t$, the future value $x(t_*)$, $t_* = h(t_0) > t_0$ is not known at time t_0 so that $f(t_0, x[h(t_0)])$ is undefined and the Euler scheme

$$x(t_0 + \Delta t) = x(t_0) + f(t_0, ?)\Delta t \tag{2.3}$$

cannot start from a given initial condition $x(t_0)$. Even if a forward function $x^0 = \varphi[h(t)]$ were somehow postulated on an interval for $t \geq t_0$, and $x(t)$ could be computed by the Euler scheme

$$x(t_0 + \Delta t) = x(t_0) + f(t_0, \varphi[h(t_0)])\Delta t, \tag{2.4}$$

this would lead to a difficulty, since for a solution $x(t)$ so obtained from (2.4), it usually happens that $x(t) \not\equiv \varphi(t)$ for $t \geq h(t_0)$, in contradiction with the requirement (2.1).

Due to such difficulties, the use of noncausal equations and inequalities was very limited, and there exists a tradition discouraging research in this area and the use of noncausal models. It is regrettable because those difficulties of solution are not in the nature of noncausal models themselves—they are rooted in the traditional time-sanctified methodology of pointwise propagation inherited from the past, but improper and inapplicable to new problems.

For many physical processes and dynamical systems, noncausal models may be of much interest. For example, the equation

$$\dot{x}(t + \alpha) - \dot{x}(t) = g\alpha, \qquad g = \text{const}, \quad \forall\, \alpha = \text{const}, \tag{2.5}$$

that can be considered as causal or noncausal depending on which derivative is taken to the right-hand side, represents a fall in the gravitational field. Indeed, differentiating (2.5) yields $\ddot{x}(t + \alpha) = \ddot{x}(t) = \text{const}, \forall\, \alpha$. The solution of (2.5) can be found by fitting with a power series which yields

$$x(t) = x_0 + v_0 t + \frac{1}{2} g t^2, \qquad x_0 = \text{const}, \quad v_0 = \text{const}. \tag{2.6}$$

If (2.5) is treated as a delay differential equation, it is easy to see that arbitrary initial function cannot be assigned since (2.5) is required to be met for any $\alpha = \text{const}$. Equation (2.5) represents such motion that differences between velocities are proportional to time lags. This property is not evident from Newtonian representation $\ddot{x}(t) = g$, nor from the explicit solution (2.6). Another model of a gravitational fall is given by the equation

$$\frac{\dot{x}(t) - t}{t - \dot{x}^{-1}(t)} = g = \text{const}, \tag{2.7}$$

which is noncausal due to the presence of the unknown inverse function $\dot{x}^{-1}(t)$, preventing explicit identification of an acting force. Equation (2.7) has the same unique solution (2.6) and the same Newtonian representation $\ddot{x}(t) = g$. It expresses the fact that in motions with constant acceleration, the skew-inverse proportion for velocities remains constant and represents the effective force/mass ratio.

Noncausal models, though presently not recognized as valid tools of scientific research, may be of much interest in cases when causal equations developed on the basis of available observations are too complicated, and the need exists to find their future implications without explicit solution of those equations. In such areas as material science, population dynamics, astrophysics, socio-economic problems, and finance, noncausal models may be of paramount importance presenting the most convenient or the only way to investigate possible future situations under incomplete or fuzzy past knowledge or on a basis of too complicated equations developed with the imposition of strong regularity conditions.

Noncausal models can be derived not only from causal equations or their solutions, they can be naturally developed through the study of experimental data. As an example, we can cite the equation

$$\dot{y}(t) = ay(\lambda t) + by(t), \qquad \lambda > 1. \tag{2.8}$$

This equation reflects actual observations [27], and its theoretical analysis can be found in [28,29]. Note that advance differential equation (2.8) can be converted into a delay differential equation

$$a\dot{y}(\tau) = \gamma\ddot{y}(\gamma\tau) - \gamma b\dot{y}(\gamma\tau), \qquad 0 < \gamma = \frac{1}{\lambda} < 1, \tag{2.9}$$

bringing extraneous solutions due to arbitrarity of initial function.

Another noncausal model is given by the equation

$$\dot{y}(t) = 2y(2t+1) - 2y(2t-1), \tag{2.10}$$

which cannot be converted into a delay differential equation. Apart from trivial solution $y(t) =$ const, this equation has nontrivial compactly supported solutions that have been studied in [30,31].

Noncausal integro-differential equations without an explicit advance term already appeared in the literature, e.g., in [32, pp. 9,17,18] in the context of a functional differential equation. The most recent occurrence, in relation to PDE, is the equation [33]

$$\begin{aligned} &\frac{\partial N}{\partial t}(t,r) + v(t,r,\alpha(t))\frac{\partial N}{\partial t}(t,r) = H(t,r,\alpha(t)), \qquad t \geq 0, \\ &\alpha(t) = \int_0^{+\infty} f(r)N(t,r)\,dr < \lambda; \qquad N(0,r) = N_0(r), \end{aligned} \tag{2.11}$$

representing a type of brittle material behavior models [34].

Models with a noncausal integral term or an explicit advance term are rarely seen in the literature, partly due to the classical tradition of causality. With time, we may expect this tradition to disappear, with the understanding that noncausal models in mathematics represent not a misconception, but a convenient formal description of some existing or assumed properties of physical processes, or motions with real forces that are not explicit as in Newton or Lagrange equations, or in the expressions of kinetic and potential energy, but are masked in noncausal terms; they are to be determined as effective forces after the solution, if it exists, of noncausal equations.

3. LOCAL OPTIMALITY CONDITIONS VERSUS GLOBAL OPTIMIZATION AND CONTROL

Traditional methods of optimization (mathematical programming), optimal control, and differential games are based on some kind of gradient or variational optimality condition [35–40] which is applied either to find the optimal solution directly or to construct an approximation process. Inherited from earlier historical developments in differential calculus and the calculus of variations, all those conditions and their generalizations [41,42] are local and they require some smoothness and convexity assumptions, the existence of certain continuously-differentiable auxiliary functions, and/or other properties commonly called regularity conditions. Local approach of consideration of a neighborhood around a candidate point and pointwise descent-ascent methods render only a local optimum, and only one corresponding local optimizer. In many cases, such optimality conditions fail altogether. Consider two simple examples.

EXAMPLE 1. Solve

$$\min x \qquad \text{or} \qquad \max x, \tag{3.1}$$

$$\text{subject to } g_1 = x \geq 0, \quad g_2 = y \geq 0, \quad g_3 = (1-x)^3 - y \geq 0. \tag{3.2}$$

The solution of this example is trivial,

$$\min x = 0, \qquad \text{for } \{x, y\} = \{0, [0,1]\}, \tag{3.3}$$

$$\max x = 1, \qquad \text{for } \{x, y\} = (1, 0). \tag{3.4}$$

However, the Karush-Kuhn-Tucker necessary conditions [35,36],

$$\nabla f\left(z^0\right) = \sum_{i=1}^{3} \lambda_i \nabla g_i\left(z^0\right), \qquad \lambda_i g_i\left(z^0\right) = 0, \tag{3.5}$$

fail in both cases. Here, ∇ denotes gradient, $z^0 = \{x^0, y^0\}$ is an optimizer, constants λ_i are nonnegative or nonpositive according to the case in (3.1), (3.2) for binding constrains, and zeros for nonbinding constraints.

If we apply to this example the necessary and sufficient global optimality condition

$$c\mu(H_c) = \int_{H_c} f(z)\, d\mu, \tag{3.6}$$

for the problem

$$\min f(z), \qquad \text{or} \qquad \max f(z), \qquad z \in Z, \tag{3.7}$$

where c is the global optimum value (a number), μ is an appropriate measure, and H_c are level sets defined as follows:

$$H_c = \{z \in Z \mid f(z) \leq c\}, \qquad c \in R, \qquad \text{for min}, \tag{3.8}$$

$$H_c = \{z \in Z \mid f(z) \geq c\}, \qquad c \in R, \qquad \text{for max}, \tag{3.9}$$

then, the solution is straightforward. Indeed, (3.6) is satisfied, if $\mu(H_c) = 0$. We have

$$\mu(H_c) = \int_0^c (1-x)^3\, dx = 0, \qquad \text{if } c = 0 \qquad (\text{for min}), \tag{3.10}$$

$$\mu(H_c) = \int_c^1 (1-x)^3\, dx = 0, \qquad \text{if } c = 1 \qquad (\text{for max}), \tag{3.11}$$

which, together with (3.1), (3.2), (3.8), and (3.9), yields solutions (3.3),(3.4). To make sure that these are the only global solutions, one has to check that (3.6) is not satisfied for other values of c.

It is easy to see that in case of inequality in (3.6), always

$$c\mu(H_c) > \int_{H_c} f(z)\, d\mu, \qquad \text{for min-problem}, \tag{3.12}$$

$$c\mu(H_c) < \int_{H_c} f(z)\, d\mu, \qquad \text{for max-problem}, \tag{3.13}$$

which leads to the monotonic set-contraction procedure in case $\mu(H_c) > 0$,

$$c_{n+1} = \frac{1}{\mu\left(H_{c_n}\right)} \int_{H_{c_n}} f(z)\, d\mu \to c, \qquad H_{c_n} \to H_c, \tag{3.14}$$

for both min and max problems. A possible difficulty of evaluating level sets H_{c_n} and corresponding integrals can be dealt with using Monte Carlo methods. The procedure is applicable to any continuous function $f(z)$ and any robust set Z, cl Z = cl int Z. This elegant method was published as early as 1978, see [43], and then applied to optimal control and differential games [44,45].

EXAMPLE 2. Consider a simple problem of stabilizing a linear oscillator

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -x_1 + u, \quad x_1(0) = x_{10}, \quad x_2(0) = x_{20}, \tag{3.15}$$

and minimizing the functional

$$J(u) = \int_0^\infty \left(ax_1^2 + bx_2^2 + cu^2\right)\, dt, \qquad (a, b, c = \text{const} > 0). \tag{3.16}$$

For the optimal control u_0 and the Bellman function

$$W(x_1, x_2) = \min_u J. \tag{3.17}$$

We have the equations

$$\frac{\partial W}{\partial x_1} x_2 + \frac{\partial W}{\partial x_2}(-x_1 + u_0) + ax_1^2 + bx_2^2 + cu_0^2 = 0, \tag{3.18}$$

$$\frac{\partial W}{\partial x_2} + 2cu_0 = 0, \tag{3.19}$$

that can be readily solved yielding the optimal control u_0 as a linear function of x_1, x_2.

If we can measure only the velocity $x_2(t)$, then there still exists the optimal regulator

$$u^* = -px_2, \qquad p = \text{const}, \tag{3.20}$$

stabilizing the system (3.15) and yielding minimum to the functional (3.16). However, the Hamilton-Jacobi-Bellman equation [38], yielding (3.18),(3.19) for the system (3.15), does not have a solution of the form (3.20). Moreover, by direct calculation, one can check that the optimality principle [38], usually accepted as a self-evident axiom, is not satisfied on trajectories of (3.15) with the optimal control (3.20). This example has been published as early as 1968, see [46], and happily forgotten. Instead, people were trying to adjust the maximum principle [37], the optimality principle [38], and the Isaacs equation [39] to problems for which they are invalid.

It is worth noting that the global optimality condition (3.6) is valid for optimization problems in functional spaces supplied with the appropriate measure. It means that from (3.6), one can obtain the maximum principle, the Hamiltonian-Jacobi-Bellman equation, and the Isaacs equation, of course, under respective regularity conditions (for the relationship between the maximum principle and the HJB-equation, see [47]). However, (3.6) is the necessary and sufficient condition that yields the full global solution, i.e., the unique global optimal value and the entire set of all global optimizers, thereby without any convexity, smoothness, or complete information conditions.

The use of certain optimality conditions presents another tradition in optimization. Though it is comfortable to have an optimality condition, it should be realized that universal methods of optimization can be developed without the use of optimality conditions and without imposing strong restrictions. To be effective, such methods should be monotonic. To deliver a global optimal solution, a method should carry out a setwise filtration or contraction of the admissible set Z in (3.7) onto the set of all global optimizers [17,18,21–26,43–45].

4. INTEGRAL CALCULUS CAN BE TAUGHT BEFORE DIFFERENTIAL CALCULUS

Setwise methodology is a way of thinking, not just a convenient method for the solution of some difficult problems. Traditional emphasis on differential calculus and ordinary differential equations with pointwise propagation methods for their numerical solution impedes such thinking and setwise intuition and imagination. Many persistent traditions in the history of mathematics can be traced to follow this habit of thinking "locally". However, the positioning of differential calculus before integral calculus happened simply by the precedence of events with the notion of derivative being developed before the integral. In fact, derivative itself can be evaluated through integration. Integrals do not become more understandable for students because of precedence of derivatives. Vice versa, if integral calculus were taught just after the theory of the limit, students might be better prepared to properly absorb the idea and inherent structure of integrals before being overloaded by a mountain of information on differential calculus. The introduction of the notion of measure is not a problem. And elements of topology can be included. After studying integral calculus, the differential calculus would be much easier. However, this is a by-product, not the objective. The main idea is that reversing the order of teaching the two subjects would help students to acquire a new form of intuition and imagination and new way of thinking by setwise images without any loss in local imagination in terms of tangent lines and planes. This is, of course, the personal opinion of the author who hopes that it is worth taking a serious look and a test.

REFERENCES

1. N.V. Azbelev, Impact of certain traditions onto development of the theory of differential equations, *Computers Math. Applic.* (this issue).
2. E.A. Galperin, Validity of feedback controls depending on higher order derivatives, *Computers Math. Applic.* **25** (10/11), 173–185 (1993).
3. S. Saks, *Theory of the Integral*, Warsaw, (1937).
4. I. Newton, *Philosophiae Naturalis Principia Mathematica*, Printed under the auspices of the London Royal Society, London, (1687); English translation, *Mathematical Principles of Natural Philosophy*, (Translated by A. Motte in 1729) University of California Press (1934) .
5. H. Lamb, *Hydrodynamics*, Sixth edition, Dover, Mineola, NY, (1945); (Originally published as *"Treatise on the Mathematical Theory of the Motion of Fluids")*, (1879).
6. E.A. Galperin, Dynamical equations with accelerations and higher order derivatives of motion in the right-hand sides are in agreement with Newton's laws, In Invited lectures delivered at the 6^{th} *International Colloquium on Differential Equations*, Plovdiv, Volume 1, pp. 171–182, (1995).
7. B. de Jager, Acceleration assisted tracking control, *IEEE Control Systems* **14** (5), 20–27 (1994).
8. S. Futami, N. Kyura and S. Mara, Vibration absorption control of industrial robots by acceleration feedback, *IEEE Trans. Ind. Electronics* **IE-30**, 299–305 (1983).
9. P.T. Kotnik, S. Yurkovich and U. Ozguner, Acceleration feedback for control of a flexible manipulator arm, *J. Robotic Systems* **5**, 181–196 (1988).
10. F. Berlin, W.M. Grimon and P.M. Frank, On robot motion control with acceleration feedback, In *Proceedings of the 2^{nd} IFAC Symposium on Robot Control, IFAC*, pp. 183–188, Pergamon Press, Oxford, (1989).
11. Y. Mori and T. Tchida, Proposal of a novel motion control method based on acceleration control, *Electr. Engrg. Japan* **110** (5), 67–76 (1990).
12. F. Matsuno and Y. Sakawa, A simple model of flexible manipulators with six axes and vibration control using accelerometers, *J. Robotic Systems* **7**, 575–597 (1990).
13. J.E. McTuroy and G.N. Saridis, Acceleration and torque feedback for robotic control: Experimental results, *J. Robotic Systems* **7**, 813–832 (1990).
14. F.R. Shaw and K. Srivivasan, Bandwidth enhancement of position measurements using measured acceleration, *Mech. Systems Sign. Proc.* **4**, 23–28 (1990).
15. E. Sim and S.W. Lee, Active vibration control of flexible structures with acceleration feedback, *J. Guidance, Control and Dynamics* **16**, 413–415 (1993).
16. P. Chiacchio, F. Pierrot and B. Siciliano, Experimenting acceleration feedback loop for robot control, In *Proceedings of the 2^{nd} European Control Conference*, Volume 2, pp. 565–569, (1993).
17. R.E. Moore, Global optimization to prescribed accuracy, *Computers Math. Applic.*, Special Issue on Global Optimization, Control and Games **21** (6/7), 25-39 (1991).
18. H. Ratshek and J. Rokne, *New Computer Methods for Global Optimization*, Ellis Horwood and John Wiley, New York, (1988).

19. R. Horst and H. Tuy, *Global Optimization, Deterministic Approaches*, Springer-Verlag, (1990).
20. A.H.G. Rinnooy Kan and G.T. Timmer, Stochastic global optimization methods, *Mathematical Programming* **39** (1), 27–78 (1987).
21. V.K. Chichinadze, *Solution of Nonconvex Nonlinear Optimization Problems (The Method of Ψ-Transform)*, Nauka (in Russian), (1983).
22. S.H. Chew and Q. Zheng, *Integral Global Optimization*, Lecture Notes in Economics and Systems, Springer-Verlag, Vol. 298, (1988).
23. E.A. Galperin, *The Cubic Algorithm for Optimization and Control*, NP Research, Montreal, Quebec, Canada, (1990).
24. E.A. Galperin, The fast cubic algorithm, *Computers Math. Applic.* **25** (10/11), 147–160 (1993).
25. E.A. Galperin, The cubic algorithm for global games with application to pursuit-evasion games, *Computers Math. Applic.* **26** (6), 13–31 (1993).
26. E.A. Galperin, Spherical algorithms and evaluation of the fixed point set, *JNA-TMA: J. of Nonlinear Analysis: Theory, Methods and Applications* **23** (12), 1545–1557 (1994).
27. V.A. Ambartsumian, On the fluctuation of brightness in galaxy, *Soviet. Math. Doklady* **44**, 223–226 (1944).
28. T. Kato and J.B. McLeod, The functional-differential equation $y'(x) = ay(\lambda x) + by(x)$, *Bull. Amer. Math. Soc.* **77** (6), 891–937 (1971).
29. T. Kato, Asymptotic behavior of solutions of the functional-differential equation $y'(x) = ay(\lambda x) + by(x)$, In *Delay and Functional Differential Equations and their Applications*, pp. 197–217, Academic Press, New York, (1972).
30. V.L. Rvachov and V.A. Rvachov, *Nonclassical Methods of Approximation Theory in Boundary-Value Problems*, Naukova Dumka, Kiev (in Russian), (1979).
31. V.L. Rvachov and V.A. Rvachov, Compactly supported solutions of functional-differential equations and their applications, *Russian Math. Surv.* **45** (1), 87–120 (1990).
32. N.V. Azbelev, V.P. Maximov and L.F. Rakhmatullina, *Introduction to the Theory of Functional Differential Equations*, Nauka, Moscow (in Russian) (1991), English translation by World Federation Publishers, U.S.A., (1995).
33. E. Garbayo, A transport functional equation, *Ann. Sci. Math. Quebec* **20** (2), 155–167 (1996).
34. D.R. Curran, L. Seaman and D.A. Shockey, Dynamic failure of solids, *Phys. Rep.* **147**, 253–388 (1987).
35. W. Karush, Minima of functions of several variables with inequalities as side conditions, Master's Thesis, University of Chicago, (1939).
36. H.W. Kuhn and A.W. Tucker, Nonlinar programming, In *Proceedings of the 2^{nd} Berkeley Symposium on Mathematical Statistics and Probability*, (Edited by J. Neyman), University of California Press, Berkeley, CA, pp. 481–492, (1951).
37. L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze and E.F. Mischenko, *The Mathematical Theory of Optimal Processes*, John Wiley & Sons, New York, (1962).
38. R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, NJ, (1957).
39. R. Isaacs, *Differential Games*, Wiley, New York, (1965).
40. G. Leitmann, *The Calculus of Variations and Optimal Control*, Plenum Press, New York, (1983).
41. R.T. Rockafellar, *The Theory of Subgradients and its Applications to Problems of Optimization*, Lecture Notes, Montreal University Press, Montreal, (1978).
42. F.H. Clarke, Generalized gradients and applications, *Trans. American Math. Soc.* **205**, 247–262 (1975).
43. Q. Zheng, B. Jiang and S. Zhuang, A method for finding the global extremum, *Acta Math. Applic. Sinica* (in Chinese) (2), 161–174 (1978).
44. E.A. Galperin and Q. Zheng, *Global Solutions in Optimal Control and Games*, NP Research, Montreal, Quebec, Canada, (1991).
45. E.A. Galperin and Q. Zheng, *New Theory of Continuous Games*, NP Research, Montreal, Quebec, Canada, (1990).
46. E.A. Galperin and E.I. Dergacheva, Optimum stabilization of a linear system with incomplete information, *Automation and Remote Control* (8), 1193–1202 (1968).
47. F.H. Clarke and R.B. Vinter, The relation between the maximum principle and dynamic programming, *SIAM Journal on Control and Optimization* **25** (5), 1291–1311 (1987).